

What is claimed is:

1. A method for classifying a plurality of objects comprising the steps of:

(a) providing one or more observations on the objects;

5 (b) assigning observations as a matrix in a multidimensional space, said matrix having at least two directions for each object of the interest in said space;

(c) identifying latent classes of each object according to a formula

$$f(\bar{Y}_{j_1, \dots, j_K}) | \{j_k \in S_{km_k}\}_{k=1}^K \sim G \left[h \left(k, j_k, \{S_{km}\}_{m=1}^{M_k} \right) \right]; \text{ and}$$

(d) calculating the likelihood that each object of the interest belongs to identified latent
10 classes.

2. The formula of claim 1 wherein $k \in \{1, \dots, K\}$ indexes the directions of the multidimensional space; $j_k \in \{1, \dots, N_k\}$ identifies an object in direction k; N_k is the number of objects in principal direction k; $\bar{Y}_{j_1, \dots, j_K}$ is a vector of one or more observations on a set of objects $\{j_1, \dots, j_K\}$; $m \in \{1, \dots, M_k\}$ indexes latent classes in direction k with M_k being the number of latent classes in direction k; S_{km} is a latent class m in direction k; $G[\cdot]$ is a specified univariate or multivariate distribution; and $f(\cdot)$ and $g(\cdot)$ are specified functions.

3. The method of claim 1 wherein said objects are classified simultaneously or sequentially.

4. A method for identifying one or more genes linked to a cellular phenotype comprising:

(a) recording in a matrix one or more measurements on each of the genes subjected to a series of experimental or observational conditions, said measurements forming a first direction in a
25 multidimensional space;

(b) providing measurements on a cell or tissue samples subjected to the essentially same series of experimental or observational conditions as in step (a), said measurements forming a second direction;

(c) identifying latent classes of the genes in the first direction, and latent classes of cell or

5 tissue samples in the second direction according to formula:

$$\log(Y_{ij}) | i \in S_m, j \in G_l \sim N \left[t_{il} + f(\alpha_{mi}, \beta_{lj}, \gamma_{ml}), \sigma^2 \right], \text{ and}$$

(d) calculating the likelihood that each gene is a member of each identified latent class for the first direction, while also calculating, simultaneously or serially, the likelihood that each cell or tissue sample is a member of each identified latent class for the second direction.

10

5. The formula according to claim 4 wherein $N[\cdot]$ refers to a Gaussian distribution; S_m is a latent class m in the first direction; G_l is a latent class l in the second direction; and $f(\alpha_{mi}, \beta_{lj}, \gamma_{ml})$ is a function of the mean parameters of a sample category, gene category, or both.

6. The method according to claim 4 wherein the cellular phenotype comprises a disease, a cellular process, a physiological pathway, a signaling pathway, a protein expression, a drug effect, or combination thereof.

7. The method according to claim 4 whereby disabilities, medications, comorbidities, laboratory results, and clinical characteristics are linked to a clinical condition in a host.

8. The method according to claim 4 whereby laboratory and observational measurements are linked to physical processes in inorganic substances.

25

9. The method according to claim 4 whereby chemical substances are linked to their respective pharmacological activities.

10. The method according to claim 4 whereby a financial performance of stocks is identified.

30

11. A method of determining in a sample a gene or cluster of genes linked to a disease using a microarray, said microarray including at least one known nucleic acid sequence, an expression and position information, comprising:

- (a) extracting expression and position information to generate a set of data corresponding to at least one dimension;
- (b) assigning in a computer to each dimension of the gene or cluster of genes a numerical value;
- (c) generating in a computer an information algorithm for said extracted information to provide a linking pattern for said gene or cluster of genes; and
- (d) determining whether the gene or cluster of genes in a sample are linked to the disease by extrapolating from the dimension-based numerical values.

12. The method according to claim 11 wherein the information algorithm is constructed in accordance with the formula $\log(Y_{ij}) | i \in S_m, j \in G_l \sim N[t_{il} \cdot f(\alpha_{mi}, \beta_{lj}, \gamma_{ml}), \sigma^2]$, wherein i and j are the expression data by gene and sample respectively; m and l are latent classes on the corresponding dimensions; the t refer to gene expression intensity parameters; and various forms for the function f are chosen.

13. A method for identifying in a library a gene or set of genes linked to metastatic properties of a cancer comprising the steps of:

- (a) providing a nucleic acid material from a suspected cancerous sample;
- (b) hybridizing the sample-derived probes to the library;
- (d) detecting the differences between hybridization results of the sample and a reference standard;
- (e) recording the differences to form a first set of data;
- (f) analyzing protein expression data to form a second set of data; and
- (g) combining said first set of data and said second set of data to identify the gene or set of genes which govern metastatic properties of the cancer.

14. A method for predicting a metastasizing potential of a cancer, comprising:

- (a) providing a tissue sample from a subject;

- (b) recording predictive parameters, wherein the predictive parameters are univariate or multivariate morphometric descriptors; and
- (c) predicting the metastasizing potential of the cancer by a statistical comparison of the recorded predictive parameters with predictive parameters of a reference sample.

5

15. The method according to claim 13, wherein the morphometric descriptors are selected from the group comprising optical density, object size, object shape, object color, amount of DNA or RNA, angular second moment, contrast, correlation, difference moment, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, maximal correlation coefficient, coefficient of variation, peak transition probability, diagonal variance, diagonal moment, second diagonal moment, product moment, triangular symmetry, sum entropy, standard deviation, cell classification (1=Hypodiploid, 2=Diploid, 3=S-Phase, 5=Tetraploid, 6=Hyperploid), blobness, perimeter, DNA index, maximum diameter, minimum diameter, elongation, run length, configurable run length and combination thereof.

15

16. A method of screening for a drug that modulates an expression of a gene or cluster of genes in a cell of interest comprising the steps of:

20

- a) exposing said cell to said drug;
- b) analyzing the gene expression in said cell, and
- c) comparing by the method of claim 4 the difference in gene expression of a drug-exposed cell to gene expression of a cell not exposed to the drug or exposed to a drug with known properties.

25

17. A method for identifying a gene or set of genes linked to a disease of interest comprising the steps of:

30

- a) registering measured observations of the gene or set of genes as variables associated with said disease at a zero time;
- b) describing the variables as a matrix in a multidimensional space, wherein each variable represents at least one first and least one second dimension in said space;
- c) carrying out, simultaneously or at later times, a series of experimental observations;

- d) determining projections of the experimental observations onto the first and second directions, whereby a multivariate model is obtained;
- e) updating during the course of the multivariate analysis at least the first and second directions of the matrix in multidimensional space, whereby the multivariate model provides the likelihood of the gene or set of genes being linked to the disease of interest.

18. The method of claim 1 wherein said method is used for identifying genes linked to cell or tissue samples collected from a host having or suspected to have a disease comprising the steps of:

- (a) assigning in a matrix one or more measurements on each of the genes over a series of experimental or observational conditions;
- (b) having genes to form a first direction in a multidimensional space;
- (c) allowing cell or tissue samples collected under differing experimental conditions to form a second direction in a multidimensional space;
- (d) identifying latent classes of genes in the first direction and latent classes of cell or tissue samples in the principal direction;
- (e) calculating the likelihood that each gene is a member of each latent class identified for the first principal direction; and
- (f) calculating a likelihood that each cell or tissue sample is a member of each latent class for the second principal direction.

19. A method for classifying a plurality of objects in an image comprising the steps of:

- a) inputting at least two distinct images of said objects;
- b) extracting a plurality of characteristics from the images, whereby at least one characteristic of the object in one image is correlated to another characteristic of the object from another image; and
- c) generating the classification result according to object membership rules, which represent a relation between the plurality of characteristics and said at least two images.

- 20. A method of generating membership rules for objects of interest by using a computer, said method comprising the steps of:

- (a) placing measurements of a first set of objects into a database of said computer, wherein members of said first set of objects, individually, do not necessarily have any hierarchical attributes or characteristics in common;
- (b) introducing measurements of a second set of objects into a second database of said computer;
- (c) generating the membership rules by including members of said first set of objects and excluding those members of said second set of objects whose individual measurements match with corresponding individual measurements of objects of said first set of objects; and
- (d) updating said membership rules by introducing measurements of additional sets of objects and adjusting matching criteria.

21. A method for identifying a gene or set of genes linked to metastatic properties of a tumor comprising the steps of:

- (a) providing a sample from suspected tumor;
- (b) extracting an experimental genetic data;
- (d) detecting the differences between extracted genetic data from the sample and a reference standard; and
- (e) generating mathematically an acceptance criteria which allows to link metastatic properties of the cancer to the gene or set of genes.

22. A method of identifying among a plurality of genes of known and unknown function those genes that are linked to a condition of interest, comprising:

- (a) providing a mathematical model which utilizes the input data to set rejection margins;
- (b) entering an experimental data from the plurality of genes of known and unknown function; and
- (c) selecting genes linked to the condition of interest based on an acceptability criteria of the mathematical model.

23. A method for analyzing an image of a plurality of objects arranged as an output signal matrix by comparing to a stored output signal matrix database which sets membership rules for the objects of interest, comprising steps:

- (a) constructing a stimulated physical matrix comprising an ordered array of objects having X and Y coordinates;
- (b) detecting the physical signal at each said object of the physical matrix;
- (c) transforming each said physical signal to generate a corresponding electrical output signal;
- 5 (d) storing each electrical output signal in an output signal matrix database associating each output signal with the X and Y coordinates of the corresponding physical matrix unit; and
- (e) determining the membership of the objects of interest by comparing the output signal matrix database of step (d) with the stored output signal matrix database.

114205.204